

STIC Biotechnology Systems Branch

RAW SEQUENCE LISTING

ERROR REPORT

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: 10/522,366  
Source: PCT  
Date Processed by STIC: 2/7/06

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.

PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:

- 1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,
- 2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY

FOR CRF SUBMISSION AND PATENTIN SOFTWARE QUESTIONS, PLEASE CONTACT MARK SPENCER, TELEPHONE: 571-272-2510; FAX: 571-273-0221

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE CHECKER VERSION 4.4.0 PROGRAM, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW FOR ADDRESS:

<http://www.uspto.gov/web/offices/pac/checker/chkrnote.htm>

Applicants submitting genetic sequence information electronically on diskette or CD-Rom should be aware that there is a possibility that the disk/CD-Rom may have been affected by treatment given to all incoming mail.

Please consider using alternate methods of submission for the disk/CD-Rom or replacement disk/CD-Rom.

Any reply including a sequence listing in electronic form should NOT be sent to the 20231 zip code address for the United States Patent and Trademark Office, and instead should be sent via the following to the indicated addresses:

1. EFS-Bio (<http://www.uspto.gov/ebc/efs/downloads/documents.htm>) , EFS Submission User Manual - ePAVE)
2. U.S. Postal Service: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450
3. Hand Carry, Federal Express, United Parcel Service, or other delivery service (EFFECTIVE 01/14/05):  
U.S. Patent and Trademark Office, Mail Stop Sequence, Customer Window, Randolph Building, 401 Dulany Street, Alexandria, VA 22314

Revised 01/10/06

BEST AVAILABLE COPY

## Raw Sequence Listing Error Summary

### ERROR DETECTED

### SUGGESTED CORRECTION

SERIAL NUMBER:

10/522,366

ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE

- 1 \_\_\_\_ Wrapped Nucleics  
    Wrapped Aminos     The number/text at the end of each line "wrapped" down to the next line. This may occur if your file was retrieved in a word processor **after** creating it. Please adjust your right margin to .3; this will prevent "wrapping."
  
- 2 \_\_\_\_ Invalid Line Length     The rules require that a line **not exceed** 72 characters in length. This includes white spaces.
  
- 3 \_\_\_\_ Misaligned Amino  
    Numbering     The numbering under each 5<sup>th</sup> amino acid is misaligned. Do **not** use tab codes between numbers; use **space characters**, instead.
  
- 4 \_\_\_\_ Non-ASCII     The submitted file was **not** saved in ASCII(DOS) text, as **required** by the Sequence Rules. **Please ensure your subsequent submission is saved in ASCII text.**
  
- 5 \_\_\_\_ Variable Length     Sequence(s) \_\_\_\_ contain n's or Xaa's representing more than one residue. **Per Sequence Rules, each n or Xaa can only represent a single residue.** Please present the **maximum** number of each residue having variable length and indicate in the <220>-<223> section that some may be missing.
  
- 6 \_\_\_\_ PatentIn 2.0  
    "bug"     A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequences(s) \_\_\_\_\_. Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence. **This applies to the mandatory <220>-<223> sections for Artificial or Unknown sequences.**
  
- 7 \_\_\_\_ Skipped Sequences  
    (OLD RULES)     Sequence(s) \_\_\_\_ missing. If intentional, please insert the following lines for **each** skipped sequence:  
                            (2) INFORMATION FOR SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)  
                            (i) SEQUENCE CHARACTERISTICS: (Do not insert any subheadings under this heading)  
                            (xi) SEQUENCE DESCRIPTION:SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)  
                            This sequence is intentionally skipped  
  
                            Please also adjust the "(ii) NUMBER OF SEQUENCES:" response to **include** the skipped sequences.
  
- 8 \_\_\_\_ Skipped Sequences  
    (NEW RULES)     Sequence(s) \_\_\_\_ missing. If **intentional**, please insert the following lines for **each** skipped sequence.  
                            <210> sequence id number  
                            <400> sequence id number  
                            000
  
- 9 \_\_\_\_ Use of n's or Xaa's  
    (NEW RULES)     Use of n's and/or Xaa's have been detected in the Sequence Listing.  
                            Per 1.823 of Sequence Rules, use of <220>-<223> is **MANDATORY** if n's or Xaa's are present.  
                            In <220> to <223> section, please explain location of **n** or **Xaa**, and which residue **n** or **Xaa** represents.
  
- 10 \_\_\_\_ Invalid <213>  
    Response     Per 1.823 of Sequence Rules, the only **valid** <213> responses are: Unknown, Artificial Sequence, or scientific name (Genus/species). <220>-<223> section is **required** when <213> response is Unknown or is Artificial Sequence
  
- 11 \_\_\_\_ Use of <220>     Use of <220> to <223> is **MANDATORY** if <213> "Organism" response is "Artificial Sequence" or "Unknown." Please explain source of genetic material in <220> to <223> section.  
                            (See "Federal Register," 06/01/1998, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of Sequence Rules)
  
- 12 \_\_\_\_ PatentIn 2.0  
    "bug"     Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing). Instead, please use "File Manager" or any other manual means to copy file to floppy disk.
  
- 13 \_\_\_\_ Misuse of n/Xaa     "**n**" can **only** represent a single nucleotide; "**Xaa**" can **only** represent a single amino acid



p4

## RAW SEQUENCE LISTING

DATE: 02/07/2006

PATENT APPLICATION: US/10/522,366

TIME: 12:49:02

Input Set : A:\Sequence\_Listing\_10522366.txt

Output Set: N:\CRF4\02012006\J522366.raw

3 &lt;110&gt; APPLICANT: National Institute of Advanced Industrial Science and Technology

5 &lt;120&gt; TITLE OF INVENTION: Lethal gene markers for transformant selection

7 &lt;130&gt; FILE REFERENCE: 332-01229

-&gt; 9 &lt;140&gt; CURRENT APPLICATION NUMBER: US/10/522,366

-&gt; 10 &lt;141&gt; CURRENT FILING DATE: 2005-01-25

12 &lt;160&gt; NUMBER OF SEQ ID NOS: 24

14 &lt;170&gt; SOFTWARE: PatentIn Ver. 2.1

Does Not Comply  
Corrected Diskette Needed

(Pg. 1-2) ↗

## RORED SEQUENCES

435 &lt;210&gt; SEQ ID NO: 20

436 &lt;211&gt; LENGTH: 330

437 &lt;212&gt; TYPE: DNA

438 &lt;213&gt; ORGANISM: E.coli

E-&gt; 440 &lt;400&gt; SEQUENCE: ①-20

441 ggccgcctcg gccgtagtag tagaaagggtt ttaaagatta cgggcatgat tatcatccag 60

442 ctccgaaaac tgagaatatt aaagggcttg gtgatcttaa gcctgggata ccaaaaacac 120

443 caaagcagaa tgggtggtgga aaacgcaagc gctggactgg agataaaggc cgtaagattt 180

444 atgagtggga ttctcagcat ggtgagcttg aggggtatcg tgccagtgat ggtcagcatc 240

445 ttggctcatt tgaccctaaa acaggcaatc agttgaaagg tccagatccg aaacgaaata 300

446 tcaagaaata tctttgaggc catagcggcc 330

449 &lt;210&gt; SEQ ID NO: 21

450 &lt;211&gt; LENGTH: 60

451 &lt;212&gt; TYPE: DNA

452 &lt;213&gt; ORGANISM: Artificial Sequence

454 &lt;220&gt; FEATURE:

455 &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: adapter

E-&gt; 457 &lt;400&gt; SEQUENCE: ②-21

458 gatccccggg taccgaggcc gcctcggccg agctcgaatt cggccggcca tagcggccgc 60

461 &lt;210&gt; SEQ ID NO: 22

462 &lt;211&gt; LENGTH: 60

463 &lt;212&gt; TYPE: DNA

464 &lt;213&gt; ORGANISM: Artificial Sequence

466 &lt;220&gt; FEATURE:

467 &lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: adapter

E-&gt; 469 &lt;400&gt; SEQUENCE: ③-22

470 aattgggcc gctatggccg gccgaattcg agctcggccg aggcggcctc ggtaccggg 60

473 &lt;210&gt; SEQ ID NO: 23

474 &lt;211&gt; LENGTH: 650

475 &lt;212&gt; TYPE: DNA

476 &lt;213&gt; ORGANISM: S.cerevisiae

E-&gt; 478 &lt;400&gt; SEQUENCE: ④-23

What is the  
source of genetic material?Invalid  
ResponseWhat is the  
source of genetic  
material?Invalid  
ResponseSee item # 11  
on error summary  
sheet. 2/7/20

## RAW SEQUENCE LISTING

DATE: 02/07/2006

PATENT APPLICATION: US/10/522,366

TIME: 12:49:02

Input Set : A:\Sequence\_Listing\_10522366.txt

Output Set: N:\CRF4\02012006\J522366.raw

```

479 ggccgcctcg gccaggatct ggtggcgaac aagcatgcga tatttgccga cttaaaaagc 60
480 tcaagtgtct caaagaaaaa ccgaagtgcg ccaagtgtct gaagaacaac tgggagtgtc 120
481 gctactctcc caaaaccaa aggtctccgc tgactagggc acatctgaca gaagtggaat 180
482 caaggctaga aagactggaa cagctatttc tactgatttt tcctcgagaa gaccttgaca 240
483 tgattttgaa aatggattct ttacaggata taaaagcatt gttaacagga ttatttgtac 300
484 aagataatgt gaataaagat gccgtcacag atagattggc ttcagtggag actgatatgc 360
485 ctctaacatt gagacagcat agaataagt cgacatcatc atcggaagag agtagtaaca 420
486 aaggtcaaag acagttgact gtatcgattg actcggcagc tcatcatgat aactccacaa 480
487 ttccgttga ttttatgcc agggatgctc ttcattggatt tgatttgtct gaagaggatg 540
488 acatgtcga tggcttgccc ttctgaaaa cggaccccaa caataatggg ttctttggcg 600
489 acggttctct cttatgtatt cttcgtgac tgactgaggc catagcggcc 650

```

492 &lt;210&gt; SEQ ID NO: 24

493 &lt;211&gt; LENGTH: 535

494 &lt;212&gt; TYPE: DNA

495 &lt;213&gt; ORGANISM: A. oryzae

E-&gt; 497 &lt;400&gt; SEQUENCE: 5-24

```

498 ggccgcctcg gccattacta gtctactagt aactctgtct tategtcatc tcccataggt 60
499 gagtttggtt gttttgtttc cactgagatc atgacctcct cctacccac catccacta 120
500 tttttgttac ggtagccatg acccctccat ggcaaagaga gaggaggacg aggacgatca 180
501 ggaaactgtg tctcgccgctc ataccacaat cgtgttatcc tgattgacat cttcttaa 240
502 atcgttgtaa ctgttcctga ctctcggta actgaaattg gatctcccca cactgcctc 300
503 taccttgtag tccgtgactg aaccatccga tcattctttt tgggtcgtcg gtgaacacaa 360
504 ccccgctgc tagtctcctt ccaacaccga tccagaattg ttttgatttt ccattccctt 420
505 cgtttatata tgtegtctct cctcccttc cgtctctttt cttcgtcct ccaagttagt 480
506 cgactgacca attccgcagc tcgtcaaaat gcctatcacc aaggccatag cggcc 535

```

The type of errors shown exist throughout the Sequence Listing. Please check subsequent Sequences for similarities.

## VERIFICATION SUMMARY

DATE: 02/07/2006

PATENT APPLICATION: US/10/522,366

TIME: 12:49:03

Input Set : A:\Sequence\_Listing\_10522366.txt

Output Set: N:\CRF4\02012006\J522366.raw

9 M:270 C: Current Application Number differs, Replaced Application Number

10 M:271 C: Current Filing Date differs, Replaced Current Filing Date

140 M:212 E: (34) Invalid or duplicate Sequence ID Number, SEQUENCE ID NOS:20 differs:1

157 M:212 E: (34) Invalid or duplicate Sequence ID Number, SEQUENCE ID NOS:21 differs:2

169 M:212 E: (34) Invalid or duplicate Sequence ID Number, SEQUENCE ID NOS:22 differs:3

178 M:212 E: (34) Invalid or duplicate Sequence ID Number, SEQUENCE ID NOS:23 differs:4

197 M:212 E: (34) Invalid or duplicate Sequence ID Number, SEQUENCE ID NOS:24 differs:5